

# Estimation of spatial max-stable models using threshold exceedances

Jean-Noel Bacro

I3M, Université Montpellier II

and

Carlo Gaetan

DAIS, Università Ca' Foscari - Venezia

## Abstract

Parametric inference for spatial max-stable processes is difficult since the related likelihoods are unavailable. A composite likelihood approach based on the bivariate distribution of block maxima has been recently proposed in the literature. However modeling block maxima is a wasteful approach provided that other information is available. Moreover an approach based on block, typically annual, maxima is unable to take into account the fact that maxima occur or not simultaneously. If time series of, say, daily data are available, then estimation procedures based on exceedances of a high threshold could mitigate such problems. In this paper we focus on two approaches for composing likelihoods based on pairs of exceedances. The first one comes from the tail approximation for bivariate distribution proposed by Ledford and Tawn (1996) when both pairs of observations exceed the fixed threshold. The second one uses the bivariate extension (Rootzén and Tajvidi, 2006) of the generalized Pareto distribution which allows to model exceedances when at least one of the components is over the threshold. The two approaches are compared through a simulation study according to different degrees of spatial dependency. Results show that both the strength of the spatial dependencies and the threshold choice play a fundamental role in determining which is the best estimating procedure.

*AMS 2000 subject classification:* primary 62G32, secondary 62M30.

*Keywords:* Composite likelihood, Extremal dependence, Generalized Pareto distribution, Spatial statistics.

# 1 Introduction

Extreme value theory for multivariate data is well established and there are a lot of results which can be used to characterize extreme values distributions (see, for instance, Resnick, 1987; Beirlant et al., 2004).

From a statistical point of view (Coles, 2001) there are mainly two approaches to fitting multivariate models: block maxima and threshold exceedances. Block maxima is the most used approach: assuming a parametric family of distributions, the model parameters are estimated on a block maxima sample. In practice, bivariate distributions of maxima have been frequently considered and there are few papers that deal with higher dimensions (Tawn, 1988, 1990; Coles and Tawn, 1991, 1994).

Recently, spatial extreme problems have received an increasing attention in the literature (see, among others, Coles (1993); Coles and Tawn (1996); Casson and Coles (1999); Buishand et al. (2008); Bel et al. (2008); Padoan et al. (2010) and for recent reviews Bacro and Gaetan (2012); Davison et al. (2012)). In particular max-stable processes (de Haan, 1984) take a prominent role in modeling the spatial dependence because they are a natural extension of the multivariate extreme value distributions for dealing with spatial data.

However, in spite of the relative large number of instances (Smith, 1990; Schlather, 2002; Kabluchko et al., 2009), parametric likelihood inference for such models is not easy because the related likelihood is unavailable or too cumbersome to compute. For this reason a methodology based on partial specification of the full likelihood, the composite likelihood (Lindsay, 1988; Varin et al., 2011), has gained popularity. For spatial extreme data, the composite likelihood (Smith and Stephenson, 2009; Padoan et al., 2010) is built by taking product of pairwise distributions. Then parameter estimates are obtained by fitting the model to block bivariate maxima vectors.

In this paper we are concerned with estimation procedures based on exceedances of a large threshold. In doing this we expect that the efficiency of the estimates can be improved in all situations where the paucity of data prevents the traditional approach based on the block maxima. For instance, rainfall data are frequently collected every day but for few years. If annual maxima are considered, we have got estimates by using few repetitions. A larger data set can be obtained by dealing with daily exceedances of a large threshold, even if constraints of stationarity can reduce the temporal window to some months per years.

Our aim is not suggesting new models for explaining the spatial behaviour of exceedances (see, for instance, Buishand et al., 2008; Turkman et al., 2010) but proposing and comparing two different approaches for composing likelihoods based on exceedances. The approaches rely on different specifications

of the bivariate distribution of the exceedances.

The first one comes from Ledford and Tawn (1996). These authors have proposed a tail approximation for the bivariate distribution valid when both components exceed a large threshold. Since the approximation is valid for values simultaneously larger than the threshold, a censored version of this approximation is proposed for pairs of data having one or two components below the threshold.

The second approach we have considered is based on a result for multivariate exceedances due to Rootzén and Tajvidi (2006). This result allows us to model exceedances over a threshold using a multivariate extension of the generalized Pareto distribution when at least one of the components is over the threshold. The result is asymptotic and leads to approximating the conditional distribution given that at least one component is over the threshold.

The paper is organized as follows. Spatial max-stable processes are presented in Section 2. The two bivariate models for the exceedances are described in Section 3 and the estimation procedure we propose is detailed in section 4. Section 5 presents a simulation study where the two threshold exceedances models are compared, according different degrees of spatial extreme dependencies. Finally, some perspectives are discussed in Section 6.

## 2 Spatial max-stable processes

When we deal with extreme values, the traditional approach considers the maxima over a number of replications using the results from the extreme value theory. This theory embraces the univariate and the multivariate case (see, for instance, Beirlant et al., 2004, for a recent account). In the case of spatial data maxima of observations over a fixed period are frequently collected on a finite subset of sites and the aim is to characterize the stochastic behavior on an uncountable set of unsampled sites. For this, multivariate models are unsuitable and their natural extensions are the max-stable processes (de Haan, 1984).

Let  $\{Z(s), s \in \mathcal{S}\}$  be a stochastic process on an index set  $\mathcal{S} \subset \mathbb{R}^d$ . We assume that  $Z_k(\cdot), 1 \leq k \leq n$ , are  $n$  independent copies of  $Z(\cdot)$ . The process  $Z(\cdot)$  is a (spatial) max-stable process if the multivariate distribution of  $p \geq 1$  sites  $Z(s_1), \dots, Z(s_p)$ , with  $s_i \in \mathcal{S}$  and  $i = 1, \dots, p$ , satisfies the so-called max-stability property, i.e. there exist sequences  $a_n(s) > 0$ ,  $b_n(s)$ ,  $n \geq 1$ ,

such that

$$\Pr \left( \max_{1 \leq k \leq n} \frac{Z_k(s_j) - b_n(s_j)}{a_n(s_j)} \leq z_j; 1 \leq j \leq p \right) = \Pr (Z(s_j) \leq z_j; 1 \leq j \leq p) \quad (1)$$

From the definition, any finite distribution of a max-stable process is an extreme value distribution. In the sequel we will denote the  $p$  dimensional finite distribution of a spatial max-stable process as  $G_{s_1, \dots, s_p}(z_1, \dots, z_p)$ .

If we suppose that we have  $n$  independent copies of a process  $Y(\cdot)$  such that

$$\lim_{n \rightarrow \infty} \Pr \left( \max_{1 \leq k \leq n} \frac{Y_k(s_j) - \beta_n(s_j)}{\alpha_n(s_j)} \leq z_j; 1 \leq j \leq p \right) = G_{s_1, \dots, s_p}(z_1, \dots, z_p) \quad (2)$$

for suitable sequences  $\alpha_n(s) > 0$ ,  $\beta_n(s)$ , we say that the distribution of  $Y(s_1), \dots, Y(s_p)$ , i.e.  $F_{s_1, \dots, s_p}$ , is in the attraction domain of the extreme value distribution  $G_{s_1, \dots, s_p}$ . This fact will be denoted by  $F_{s_1, \dots, s_p} \in \mathcal{D}(G_{s_1, \dots, s_p})$ .

There are several representations for max-stable processes, see for instance Smith (1990); Schlather (2002); Kabluchko et al. (2009); Lantuéjoul et al. (2011). Just for illustration, in the sequel we will concentrate on the Gaussian extreme value process (Smith, 1990).

Let  $\Pi$  be a Poisson process of intensity  $x^{-2} dx ds$  on  $(0, \infty) \times \mathcal{S}$ . Then,

$$Z(s) = \max_{(x, u) \in \Pi} x f(s - u), \quad (3)$$

defines a stationary max-stable process, with unit Fréchet marginal distribution, i.e.  $\Pr(Z(s) \leq z) = \exp(-1/z)$ ,  $z > 0$ , provided that  $f(\cdot)$  is a non negative function such that  $\int_{\mathcal{S}} f(s) ds = 1$ . Choosing  $f$  as the bivariate Gaussian density

$$f(s) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp \left( -\frac{1}{2} s' \Sigma^{-1} s \right),$$

where  $\Sigma$  is a covariance matrix, leads to the Gaussian extreme value process. Smith (1990) gives an intuitive physical interpretation of the process (3) in terms of rainfall-storms i.e.  $x$  and  $u$  are, respectively, the size and the center of a 'storm' that spreads out according the density  $f$ .

For max-stable processes analytical expressions for multivariate distributions are quite cumbersome and explicit formulas are known mainly for the bivariate case. Smith (1990) derived the bivariate distribution of (3) for two different sites  $s_i$  and  $s_j$ , namely

$$G_{s_i, s_j}(z_i, z_j) = \exp \left\{ -\frac{1}{z_i} \Phi \left( \frac{a_{ij}}{2} + \frac{1}{a_{ij}} \log \frac{z_j}{z_i} \right) - \frac{1}{z_j} \Phi \left( \frac{a_{ij}}{2} + \frac{1}{a_{ij}} \log \frac{z_i}{z_j} \right) \right\} \quad (4)$$

where  $\Phi(\cdot)$  denotes the standard normal cumulative distribution function and  $a_{ij} = \sqrt{(s_i - s_j)' \Sigma^{-1} (s_i - s_j)}$ . In the case of a marginal Gumbel distribution, i.e.  $\Pr(Z(s) \leq z) = \exp\{-\exp(-z)\}$ , the bivariate distribution is given by

$$G_{s_i, s_j}(z_i, z_j) = \exp \left\{ -e^{-z_i} \Phi \left( \frac{a_{ij}}{2} + \frac{z_i - z_j}{a_{ij}} \right) - e^{-z_j} \Phi \left( \frac{a_{ij}}{2} + \frac{z_j - z_i}{a_{ij}} \right) \right\}$$

More recently Genton et al. (2011) have derived multivariate expression for the Gaussian extreme value process but in practice only the three-variate case is simple to deal with.

This fact has restricted the application of the parametric inference based on the full likelihood and motivated approaches based on the composite likelihood (Lindsay, 1988). Roughly speaking (see section 4 for more details) the composite likelihood is an estimating function obtained combining marginal or conditional densities.

However, these new advances still require a lot of temporal information, because we need to calculate the maxima over several period for providing enough sample data. This weakness can be overcome if we use all observations that exceed some fixed (high) threshold. For this reason in the next section we will illustrate two different ways for modeling bivariate threshold exceedances and we will exploit these representations for building suitable composite likelihoods.

### 3 Models for bivariate threshold exceedances

Univariate extreme values results (Pickands, 1975; Davison and Smith, 1990) justify to model the conditional distribution of the excesses  $Y(s) - u$  through the Generalized Pareto (GP) distribution

$$\Pr \left( \frac{Y(s) - u}{\sigma_u} \leq z | Y(s) > u \right) \approx 1 - (1 + \xi z)^{-1/\xi}, \quad z \geq 0,$$

provided that  $F_s(\cdot) \in \mathcal{D}(G)$ . Here  $\xi$  is a real shape parameter and  $\sigma_u$  is a positive scale parameter, which depends on the threshold  $u$ .

The GP distribution provides an approximation of the tail of the distribution function in the sense that for a large enough value of  $u$ , we have

$$\Pr \left( \frac{Y(s) - u}{\sigma_u} \leq z \right) \approx 1 - \zeta \{1 + \xi z\}^{-1/\xi}, \quad z \geq 0, \quad (5)$$

where  $\zeta = 1 - F_s(u)$ .

Now we consider the bivariate random variable  $Y(s_i), Y(s_j)$  with the same marginal distributions and  $F_{s_i, s_j} \in \mathcal{D}(G_{s_i, s_j})$ . According (5), the integral transformation

$$\tilde{Y}_k = - \left( \log \left\{ 1 - \zeta \left[ 1 + \xi \frac{Y(s_k) - u}{\sigma_u} \right]^{-1/\xi} \right\} \right)^{-1}, \quad k = i, j \quad (6)$$

has unit Fréchet distribution.

The bivariate distribution  $F_{s_i, s_j}(y_i, y_j)$  can be approximated on  $(u, \infty)^2$  (Ledford and Tawn, 1996) as

$$F_{s_i, s_j}(y_i, y_j) \approx G_{s_i, s_j}(\tilde{y}_i, \tilde{y}_j). \quad (7)$$

Formula (7) gives us an approximation of the bivariate distribution in the tail when both components are greater than threshold  $u$ .

Another approach (Beirlant et al., 2004; Rootzén and Tajvidi, 2006) considers a bivariate distribution of large values, given that at least one of the components is large. This distribution is a bivariate analogue for the GP distribution.

Provided that  $F_{s_i, s_j}(\cdot, \cdot) \in \mathcal{D}(G_{s_i, s_j}(\cdot, \cdot))$  and  $0 < G_{s_i, s_j}(0, 0) < 1$ , Rootzén and Tajvidi (2006) show that for large  $u$

$$\Pr \left( \frac{Y(s_i) - u}{\sigma_u} \leq z_i, \frac{Y(s_j) - u}{\sigma_u} \leq z_j \mid Y(s_i) > u \text{ or } Y(s_j) > u \right) \approx H_{s_i, s_j}(z_i, z_j),$$

where

$$H(x_i, x_j) = \frac{1}{-\log(G_{s_i, s_j}(0, 0))} \log \frac{G_{s_i, s_j}(x_i, x_j)}{G_{s_i, s_j}(\min(x_i, 0), \min(x_j, 0))}. \quad (8)$$

Actually the Rootzen-Tajvidi's result applies to a dimension greater than two, and in the sequel we refer to the distribution (8) as the bivariate GP distribution.

## 4 Inference procedures for spatial high threshold excesses

As mentioned in section 2, likelihood inference for spatial max-stable processes is difficult because we don't know an analytical expression for the multivariate distribution of the majority of the spatial max-stable processes.

The composite likelihood (CL) (Lindsay, 1988) is an inference function built by multiplying the likelihood of marginal or conditional events. More

precisely, let  $\{f(\cdot, \theta), \theta \in \Theta\}$  be a parametric family of joint densities for the observations  $y_1, \dots, y_n \in D \subseteq \mathbb{R}^n$  and consider a set of events  $\{A_i : A_i \subseteq \mathfrak{F}, i \in I \subseteq \mathbb{N}\}$ , where  $\mathfrak{F}$  is a  $\sigma$ -algebra on  $D$ . The logarithm of the CL is defined as (Lindsay, 1988):

$$cl(\theta) = \sum_{i \in I} w_i \log f((y_1, \dots, y_n)' \in A_i, \theta),$$

where  $w_i$  are non negative weights to be fixed.

For max-stable processes, Padoan et al. (2010) have proposed a CL approach based on the pairwise distributions for estimating the parameters of the distribution of random vectors of block maxima. Their estimation procedure has been implemented in the R package **SpatialExtremes** (Ribatet, 2011) available on the CRAN repositories. Here we borrow the same idea but we use exceedances instead of block maxima.

We assume that data  $\{y_{t1}, \dots, y_{tp}\}$ ,  $t = 1, \dots, T$ , are  $T$  independent realizations of the vector  $\{Y(s_1), \dots, Y(s_p)\}$ , where  $s_1, \dots, s_p$  are  $p$  sites belonging to  $\mathcal{S}$ .

Let  $f_{s_i, s_j}(\cdot, \cdot; \theta)$  be the bivariate density associated to the distribution (7) or (8). The CL estimate  $\hat{\theta}_T$  is obtained maximizing

$$cl_T(\theta) = \sum_{t=1}^T \sum_{i=1}^p \sum_{j>i}^p w_{ij} \log f_{s_i, s_j}(y_{ti}, y_{tj}; \theta) = \sum_{t=1}^T U_t(\theta) \quad (9)$$

where  $w_{ij} \geq 0$  designate user-specified weights (Padoan et al., 2010) and  $U_t(\theta) = \sum_{i=1}^p \sum_{j>i}^p w_{ij} \log f_{s_i, s_j}(Y_{ti}, Y_{tj}; \theta)$ .

Note that there is a fundamental difference between the estimating functions, when we choose (7) or (8). In using (7) we exploit a tail approximation, moreover the approximation is sound on the region  $(u, \infty)^2$  and does not apply to pairs of observations outside that region. For pairs such that at least one component does not exceed  $u$ , we use the censored approach as proposed in Ledford and Tawn (1996). The likelihood contribution of each pair is

$$f_{s_i, s_j}(y_{ti}, y_{tj}; \theta) = \begin{cases} \nabla_{ab} G_{s_i, s_j}(\tilde{y}_{ti}, \tilde{y}_{tj}; \theta) & \text{if } y_{ti} \geq u, y_{tj} \geq u \\ \nabla_a G_{s_i, s_j}(\tilde{y}_{ti}, \tilde{u}; \theta) & \text{if } y_{ti} \geq u, y_{tj} < u \\ \nabla_b G_{s_i, s_j}(\tilde{u}, \tilde{y}_{tj}; \theta) & \text{if } y_{ti} < u, y_{tj} \geq u \\ G_{s_i, s_j}(\tilde{u}, \tilde{u}; \theta) & \text{if } y_{ti} < u, y_{tj} < u \end{cases}$$

where  $\theta = (\sigma_u, \xi, \beta)$  and  $\beta$  is the (possibly) vector of spatial dependence parameters. Instead of maximizing  $cl(\theta)$  with respect to all parameters, we

can estimate  $\theta$  in two steps. First of all we estimate the marginal parameters  $\xi$  et  $\sigma_u$ , then their estimates,  $\hat{\xi}$  and  $\hat{\sigma}_u$ , are plugged in (6) for obtaining an estimate for  $\beta$  maximizing  $cl(\hat{\xi}, \hat{\sigma}_u, \beta)$ .

When we use (8), differently we keep the pairs such that there is at least one exceedance and the number of pairs in (9) is a random variable. Moreover we refer to conditional events. The associated density is given by

$$f_{s_i, s_j}(y_{ti}, y_{tj}; \theta) = -\frac{1}{\log G_{s_i, s_j}(u, u; \theta)} \times \left[ \frac{\nabla_{ab} G_{s_i, s_j}(y_{ti}, y_{tj}; \theta) - \nabla_a G_{s_i, s_j}(y_{ti}, y_{tj}; \theta) \nabla_b G_{s_i, s_j}(y_{ti}, y_{tj}; \theta)}{G_{s_i, s_j}(y_{ti}, y_{tj}; \theta)^2} \right]$$

for  $(y_{ti}, y_{tj}) \notin (-\infty, u]^2$ . In this case the components of the parameter vector  $\theta$  are the marginal parameter  $\sigma_u$  and the dependence parameters  $\beta$ .

Exceedances above  $u$  will appear in spatial clusters and in this respect the weights  $w_{ij}$  in (9) should handle pairwise dependencies inside and outside of the clusters.

Composite log-likelihood, as linear combination of log-likelihoods, allows to obtain unbiased estimating equations under classical regularity conditions (Heyde, 1997). So, in principle, it would be possible to come up with an optimal way of combining the individual scores  $\nabla_{\theta} \log f_{s_i, s_j}(y_{ti}, y_{tj}; \theta)$  and determining optimal weights  $w_{ij}$  (Kuk, 2007). In the next section, more simply, we look at weights for reducing the computational burden as in Padoan et al. (2010), like as the cut-off weights, namely  $w_{ij} = 1$  if  $\|s_i - s_j\| \leq \delta$ , and 0 otherwise, for a fixed  $\delta > 0$ . There is some evidence for Gaussian random processes (Davis and Yau, 2011; Bevilacqua et al., 2012) that this choice improves not only the computational efficiency but also the statistical efficiency.

Under suitable conditions (see the Appendix in Padoan et al., 2010) the maximum composite likelihood estimator for  $\theta$  can be proved consistent and asymptotically Gaussian, for large  $T$ . The asymptotic variance is given by the inverse of the Godambe information matrix

$$\mathcal{G}_T(\theta) = \mathcal{H}_T(\theta) [\mathcal{J}_T(\theta)]^{-1} \mathcal{H}_T(\theta), \quad (10)$$

where  $\mathcal{H}_T(\theta) = \mathbb{E}(-\nabla^2 cl_T(\theta))$  and  $\mathcal{J}_T(\theta) = \mathbb{V}ar(\nabla cl_T(\theta))$ .

Standard error evaluation requires consistent estimation of the matrices  $\mathcal{H}_T$  and  $\mathcal{J}_T$ . Assuming strong-stationarity in time, we get

$$\mathcal{H}_T(\theta) = -T \mathbb{E} \{ \nabla^2 U_1(\theta) \},$$



that can be estimated by  $H_T = -\nabla^2 cl_T(\hat{\theta}_T)$ . Estimation of the matrix

$$\mathcal{J}_T(\theta) = \sum_{t=1}^T \sum_{k>t}^T \text{Cov} \{ \nabla U_t(\theta) \nabla U_k(\theta)' \}$$

requires some care. When we deal with real data, for instance environmental data, exceedances are seldom independent in time. In such case declustering techniques (Nadarajah, 2001) could be used to overcome the dependence in time. An alternative way is employing a subsampling technique (Carlstein, 1986) to estimate  $\mathcal{J}_T$ . The subsampling method consists of estimating  $\mathcal{J}_T$  over  $M$  overlapping temporal windows  $D_j \subset \{1, \dots, T\}$ ,  $j = 1, \dots, M$ , of size  $d_j$  by using the expression

$$J_T = \frac{T}{M} \sum_{j=1}^M \frac{\nabla cl_{D_j}(\hat{\theta}_T) \nabla cl_{D_j}(\hat{\theta}_T)'}{d_j},$$

where  $\nabla cl_{D_j}$  is the composite likelihood score evaluated over the window  $D_j$ . An estimate of the asymptotic variance is then given by

$$V_T = H_T^{-1} J_T H_T^{-1}.$$

## 5 Numerical results

In this section we describe a simulation study with the aim of examining and comparing the performances of CL estimates for the parameters of spatial max-stable processes. Three versions of the CL are contrasted: two based on threshold exceedances that use the bivariate distributions (7) and (8), noted LT and RT, respectively, and one based on block maxima data (Padoan et al., 2010), PRS hereafter. The advantage of threshold methods over the block maxima one is well known even if this advantage deserves to be evaluated (Madsen et al., 1997).

We have considered the Gaussian extreme value process (3) and we have set  $\Sigma = \begin{pmatrix} 200 & 150 \\ 150 & 300 \end{pmatrix}$  as in Padoan et al. (2010). The spatial process is observed on  $n^2$  points located over a  $n \times n$  regular grid  $\{k, \dots, n * k\}^2$ , where  $n = 5, 7$ . Here we present only the results for the 49 sites, because evidences are virtually the same with the smallest grid (25 sites). Three lags  $k$  for the spatial sites, namely  $k = 1, 5, 10$ , are considered in order to take into account different levels of the spatial extremal dependence.

The spatial extremal dependence of the maxima can be characterized using the extremal coefficient function (Smith, 1990; Schlather and Tawn, 2003),  $v(\cdot)$ , given by

$$\Pr(\max(Z(s), Z(s+h)) \leq z) = \exp\{-v(h)/z\}.$$

For any vector  $h$ , we have  $1 \leq v(h) \leq 2$  and  $v(h) = 1$  corresponds to the perfect dependence, instead for independent maxima we have  $v(h) = 2$ . In the case of the Gaussian extreme value process, the extremal coefficient function is

$$v(h) = 2\Phi\left(\frac{\sqrt{h'\Sigma^{-1}h}}{2}\right).$$

Figure 1 shows the values of this function for the grids over 49 spatial sites and for the different spatial lags. As expected, the spatial extremal dependence decreases as the spatial lag increases. For  $k = 1$ , the extremal dependence keeps strong on the whole grid. For  $k = 10$ , the dependence appears strong only for neighbouring sites.

In our experiments, for each simulation, we have considered 40 years and for each year we have a sample of 91 independent daily observations. Our aim is to mimic a framework where maxima are taken over a fixed season of the year. Then we have considered 500 simulations for evaluating mean, standard deviation, and root mean square error of the estimates.

Simulations have been carried out using the `rmaxstab` function of the R package **SpatialExtremes** (Ribatet, 2011). This function allows to simulate a finite realization  $(Z(s_1), \dots, Z(s_p))'$  of a Gaussian extreme value process with unit Fréchet marginal distributions. In order to fulfill the condition  $0 < G_{s_i, s_j}(0, 0) < 1$  of Theorem 2.1 in Rootzén and Tajvidi (2006), we have transformed the data to the Gumbel scale, namely  $Z^*(s_i) = \log(Z(s_i))$ ,  $i = 1, \dots, n$ . The threshold  $u$  for all sites has been set equal to the empirical 0.98-quantile, calculated over all data in each Monte Carlo simulation.

As mentioned at the end of Section 4, a careful choice of the weights in (9) could improve the statistical efficiency in addition to the computational one. In fitting the models, we have considered three sets of weights, according to different values of  $\delta$ , where  $\delta$  has been set equal to the  $a$ -quantile,  $q_a$ , of the distances among all pairs of sites and  $a = 0.25, 0.50, 1.00$ .

Figures (2), (3) and (4) contrast the variability of the estimates for grids with different spatial lags and different set of weights derived from  $\delta = q_{0.25}$  and  $\delta = q_{1.00}$ . The RT estimates always display the smallest variability but raise some concerns according different degrees of spatial dependence. When we use all the pairs ( $\delta = q_{1.00}$ ) in forming the CL function, the bias of the  $\sigma_{11}$  and  $\sigma_{22}$  estimates seriously increases in case of weak spatial dependency

(spatial lag  $k = 10$ ). On the other hand LT results point out an efficiency similar to the PRS method and the bias of the estimates keep within the reasonable bounds, regardless of the spatial dependence and the number of the pairs.

In Tables 1, 2 and 3 we measure the efficiency of the estimates in terms of the root mean square errors (RMSEs) under different extremal dependence and different amount of pairs. Bold figures highlight the minimum value of RMSEs. When the extremal dependence is strong (grids with lag  $k = 1$ ), the RT method clearly outperforms the LT one for whichever value of  $\delta$ . For moderate spatial dependencies (grids with lags  $k = 5$ ), there is again an advantage in preferring the RT method with respect to the LT. Results for weak spatial dependencies (grid with lag  $k = 10$ ) are quite different and put forwards the role of the number of pairs: for  $\delta = q_{0.25}$ , RT approach gives smallest RMSE than LT and for  $\delta = q_{1.00}$ , the converse happens. For  $\delta = q_{0.50}$ , there is not clear indication of which method behaves better.

However our results, consistent to the literature on weighted composite likelihood (Joe and Lee, 2009; Davis and Yau, 2011; Bevilacqua et al., 2012), display that including many pairs in the pairwise likelihood can harm the CL estimator, suggesting that we should retain as few pairs as possible.

We have already remarked that the RT estimates always had smaller standard errors than the LT ones. Nevertheless, they also lead to a larger RMSEs when the spatial dependence is weak and we use all the pairs because a large bias is present. This observed bias in the RT estimates seems also depend on the choice of the threshold. As matter of proof, we did the same Monte Carlo experiment setting the threshold  $u$  equal to the empirical 0.90, 0.95 and 0.98 quantiles of the data. These values correspond to the usual choice in the real data analysis. Table 4 reports the bias of the parameter estimates according the three values of the threshold. We can see that for lower thresholds, the absolute value of the bias increases for the  $\sigma_{11}$  and  $\sigma_{22}$  estimates following the RT approach. On the other hand the bias of the LT estimates is always moderate.

## 6 Discussion

In this paper, we have compared two threshold exceedances models to estimate the parameters of a particular instance of spatial max-stable processes, namely the Gaussian extreme value process (Smith, 1990; Padoan et al., 2010) of spatial max-stable processes. Each model is based on a particular bivariate extreme value property (Rootzén and Tajvidi, 2006; Ledford and Tawn, 1996). However our estimation strategy can be easily extended to

other spatial max-stable processes (Schlather, 2002; Kabluchko et al., 2009), provided that the expression of the bivariate distribution is easy to evaluate.

Our simulation experiments point out that both methods are valuable and choosing between the RT and LT approach relies on the degree of the spatial dependence. When the spatial dependence between high values is strong, the RT approach seems preferable. The RT approach for weak spatial dependent data displays a significative bias in the estimation. On the other hand LT approach does not suffer from this problem.

A possible explanation of this evidence is that the bivariate distribution (8) deals with pairs of observations with at least one component exceeding the threshold  $u$ . As Rootzén and Tajvidi (2006) have underlined in their paper, the multivariate generalized Pareto is degenerate in the case of independence, so it is not surprising that the weak extremal dependence can spoil the CL estimates. In this regard, our findings are consistent with the results obtained by Rakonczai and Tajvidi (2010) for a bivariate logistic extreme value with weak dependence. Moreover, the distribution (8) is not the true conditional bivariate distribution but only its asymptotic approximation. Again in the case of weak extremal dependence, the convergence may be slow, leading to worry regarding to the usual behaviour of the CL estimates.

Nevertheless our simulations highlight that a clever choice for the weights of the CL could significantly improve the efficiency of the estimates. This is especially true under weak spatial dependency, where the RT approach gives the best RMSE results. Indeed, when we consider all the pairs, pairs of sites which are not spatially dependent could be linked because of an exceedance of one component. In the bivariate CL framework, this should lead to a misleading link in the extremal association and, as a consequence, contribute to an incorrect estimate.

This fact entails that there is a effective need of simple practical rules for fixing the weights. However rules based only on the extremal coefficient function appear as not really adapted and more theoretical work seems necessary. A possible solution that we will explore in the future is to choose the weights by minimizing a certain norm of the Godambe information matrix (10) as in Bevilacqua et al. (2012).

## References

- Bacro, J. N., and Gaetan, C. (2012), “A review on spatial extreme modelling,” in *Advances and Challenges in Space-Time Modelling of Natural Events*, eds. E. Porcu, J. M. Montero, and M. Schlather, New York: Springer, pp. 103–124.

- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, New York: John Wiley & Sons.
- Bel, L., Bacro, J. N., and Lantuéjoul, C. (2008), “Assessing extremal dependence of environmental spatial fields,” *Environmetrics*, 19, 163–182.
- Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012), “Estimating space and space-time covariance functions: a weighted composite likelihood approach,” *Journal of the American Statistical Association*, to appear.
- Buishand, T. A., de Haan, L., and Zhou, C. (2008), “On spatial extremes: with application to a rainfall problem,” *The Annals of Applied Statistics*, 2, 624–642.
- Carlstein, A. (1986), “The use of subseries values for estimating the variance of a general statistic from a stationary sequence,” *The Annals of Statistics*, 14, 1171–1179.
- Casson, E., and Coles, S. G. (1999), “Spatial regression models for extremes,” *Extremes*, 1, 449–468.
- Coles, S. G. (1993), “Regional modelling of extreme storms via max-stable processes,” *Journal of the Royal Statistical Society, Series B*, 55, 797–816.
- Coles, S. G. (2001), *An introduction to statistical modeling of extreme value*, Boca Raton: FL: Springer.
- Coles, S. G., and Tawn, J. A. (1991), “Modelling extreme multivariate events,” *Journal of Royal Statistical Society, Series B*, 53, 377–392.
- Coles, S. G., and Tawn, J. A. (1994), “Statistical methods models for multivariate extremes: an application to structural design,” *Applied Statistics*, 43, 1–48.
- Coles, S. G., and Tawn, J. A. (1996), “Modelling extremes of the areal rainfall process,” *Journal of Royal Statistical Society, Series B*, 58, 329–347.
- Davis, R., and Yau, C.-Y. (2011), “Comments on pairwise likelihood in time series models,” *Statistica Sinica*, 21, 255–277.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), “Statistical modelling of spatial extremes,” *Statistical Science*, to appear.
- Davison, A. C., and Smith, R. L. (1990), “Models for exceedances over high thresholds,” *Journal of Royal Statistical Society, Series B*, 3, 393–442.

- de Haan, L. (1984), “A spectral representation for max-stable processes,” *The Annals of Probability*, 12, 1194–1204.
- Genton, M. G., Ma, Y., and Sang, H. (2011), “On the likelihood function of Gaussian max-stable processes,” *Biometrika*, 98, 481–488.
- Heyde, C. (1997), *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*, New York: Springer.
- Joe, H., and Lee, Y. (2009), “On weighting of bivariate margins in pairwise likelihood,” *Journal of Multivariate Analysis*, 100, 670–685.
- Kabluchko, Z., Schlather, M., and de Haan, L. (2009), “Stationary max-stable fields associated to negative definite functions,” *The Annals of Probability*, 37, 2042–2065.
- Kuk, A. (2007), “A hybrid pairwise likelihood method,” *Biometrika*, 94, 939–952.
- Lantu  joul, C., Bacro, J.-N., and Bel, L. (2011), “Storm processes and stochastic geometry,” *Extremes*, 14, 413–428.
- Ledford, A. W., and Tawn, J. A. (1996), “Statistics for near independence in multivariate extreme values,” *Biometrika*, 83, 169–187.
- Lindsay, B. (1988), “Composite likelihood methods,” *Contemporary Mathematics*, 80, 221–239.
- Madsen, H., Rasmussen, P. F., and Rosbjerg, D. (1997), “Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. at-site modeling,” *Water Resources Research*, 33, 747–757.
- Nadarajah, S. (2001), “Multivariate declustering techniques,” *Environmetrics*, 12, 357–365.
- Padoan, S. A., Ribatet, M., and Sisson, S. (2010), “Likelihood-based inference for max-stable processes,” *Journal of the American Statistical Association*, 105, 263–277.
- Pickands, J. (1975), “Statistical inference using extreme order statistics,” *The Annals of Statistics*, 3, 119–131.
- Rakonczai, P., and Tajvidi, N. (2010), “On prediction on bivariate extremes,” *International Journal of Intelligent Technologies and Applied Statistics*, 3, 115–139.

- Resnick, S. (1987), *Extreme values, Regular variation and Point Processes*, New York: Springer.
- Ribatet, M. (2011), *SpatialExtremes: Modelling Spatial Extremes*. R package version 1.8-1.
- Rootzén, H., and Tajvidi, N. (2006), “Multivariate generalized Pareto distributions,” *Bernoulli*, 12, 917–930.
- Schlather, M. (2002), “Models for stationary max-stable random fields,” *Extremes*, 5, 33–44.
- Schlather, M., and Tawn, J. A. (2003), “A dependence measure for multivariate and spatial extreme values: properties and inference,” *Biometrika*, 90, 139–156.
- Smith, E. L., and Stephenson, A. G. (2009), “An extended Gaussian max-stable process model for spatial extremes,” *Journal of Statistical Planning and Inference*, 139, 1266–1275.
- Smith, R. L. (1990), “Max-stable processes and spatial extremes,” Preprint. University of Surrey.
- Tawn, J. (1988), “Bivariate extreme value theory: models and estimation,” *Biometrika*, 75, 397–415.
- Tawn, J. (1990), “Modelling multivariate extreme value distributions,” *Biometrika*, 77, 245–253.
- Turkman, K., Amaral Turkman, M., and Pereira, J. (2010), “Asymptotic models and inference for extremes of spatio-temporal data,” *Extremes*, 13, 375–397.
- Varin, C., Reid, N., and Firth, D. (2011), “An overview of composite likelihood methods,” *Statistica Sinica*, 21, 5–42.

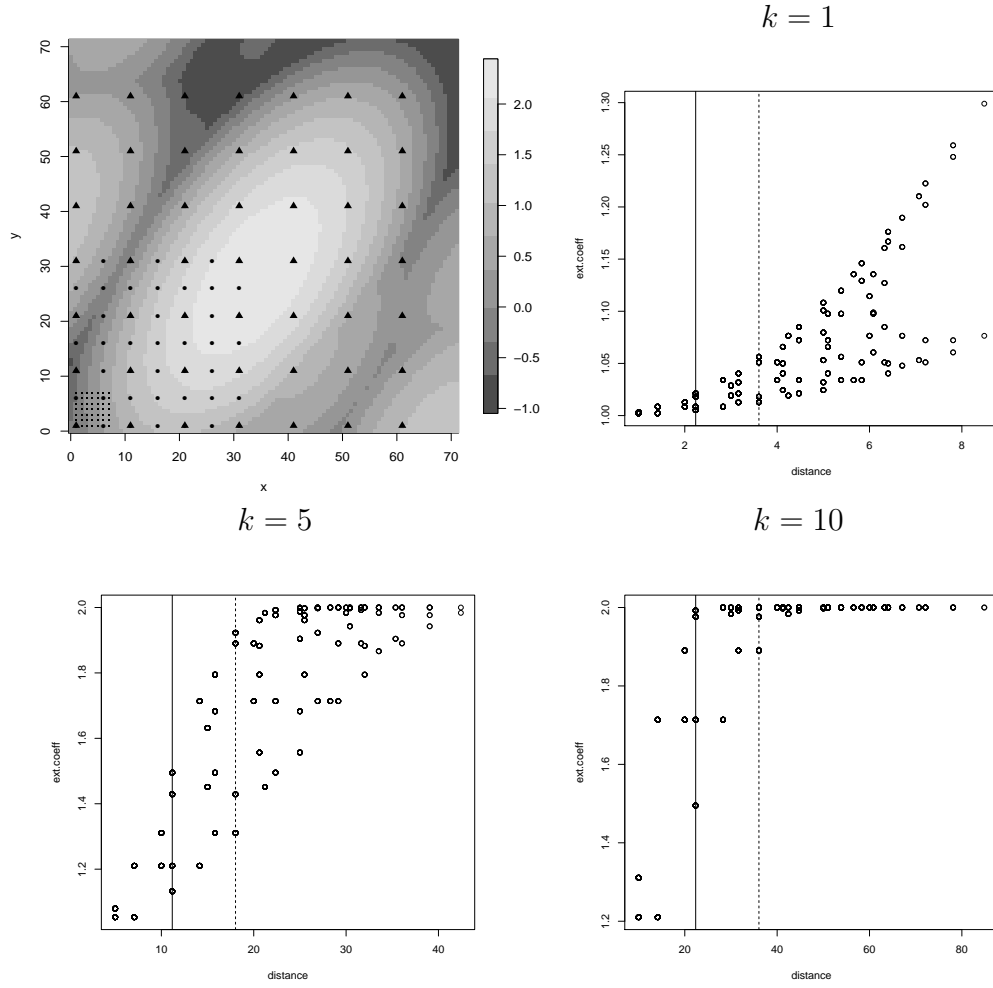


Figure 1: On the left top corner: an example of a simulation from a Gaussian extreme value process, with  $\sigma_{11} = 200$ ,  $\sigma_{22} = 300$ ,  $\sigma_{12} = 300$ . Three different regular grids, with different spatial lags ( $k = 1, 5, 10$ ), are superimposed. The other figures are the plots of the extremal coefficient function  $v(h)$ . This function is evaluated at the distances among 49 sites on regular grids with different spatial lags. Solid lines (dashed lines) correspond to the value  $\delta = q_{0.25}$  ( $\delta = q_{0.50}$ , respectively).



	Parameter	$\sigma_{11}$	$\sigma_{22}$	$\sigma_{12}$
	True value	200.00	300.00	150.00
$\delta = q_{0.25}$				
	RT	209.13 (29.62) <b>30.99</b>	308.71 (44.14) <b>44.99</b>	156.49 (30.06) <b>30.75</b>
	LT	204.24 (54.85) 55.01	301.80 (80.63) 80.65	152.71 (46.81) 46.89
	PRS	216.94 (47.77) 50.69	320.01 (67.19) 70.10	162.03 (46.94) 48.46
$\delta = q_{0.50}$				
	RT	209.00 (29.18) <b>30.54</b>	308.46 (43.53) <b>44.34</b>	156.31 (29.57) <b>30.24</b>
	LT	204.75 (54.35) 54.56	302.42 (79.73) 79.76	153.08 (46.53) 46.64
	PRS	217.53 (49.03) 52.07	320.57 (68.79) 71.80	162.36 (48.16) 49.73
$\delta = q_{1.00}$				
	RT	208.79 (28.09) <b>29.44</b>	308.29 (42.12) <b>42.92</b>	156.02 (28.45) <b>29.07</b>
	LT	204.88 (54.63) 54.85	302.43 (79.90) 79.93	153.13 (46.79) 46.89
	PRS	218.66 (51.33) 54.61	321.70 (71.49) 74.71	163.00 (50.13) 51.79

Table 1: Simulation results for a grid of 49 sites with spatial lag  $k = 1$ . For each estimation approach (RT, LT, PRS), the mean and the standard errors (in brackets) of the estimates are reported on the first line. RMSE values are given on the second line.

	Parameter	$\sigma_{11}$	$\sigma_{22}$	$\sigma_{12}$
	True value	200.00	300.00	150.00
$\delta = q_{0.25}$				
	RT	204.41 (16.33) <b>16.91</b>	305.37 (25.63) <b>26.19</b>	152.06 (16.92) <b>17.04</b>
	LT	201.06 (35.39) 35.41	302.03 (56.71) 56.74	151.55 (32.59) 32.63
	PRS	206.78 (29.79) 30.55	312.02 (49.35) 50.79	155.90 (31.49) 32.04
$\delta = q_{0.50}$				
	RT	205.58 (15.54) <b>16.51</b>	304.47 (24.01) <b>24.42</b>	148.33 (15.89) <b>15.97</b>
	LT	201.53 (35.92) 35.95	302.52 (57.17) 57.23	151.87 (33.07) 33.13
	PRS	208.22 (33.67) 34.66	313.99 (54.34) 56.11	156.79 (35.34) 35.99
$\delta = q_{1.00}$				
	RT	216.32 (13.80) <b>21.37</b>	306.31 (19.81) <b>20.79</b>	133.11 (13.26) <b>21.47</b>
	LT	202.20 (37.21) 37.27	303.47 (58.34) 58.44	152.31 (33.65) 33.73
	PRS	211.34 (41.89) 43.40	318.21 (64.93) 67.43	158.78 (42.98) 43.87

Table 2: Simulation results for a grid of 49 sites with spatial lag  $k = 5$ . For each estimation approach (RT, LT, PRS), the mean and the standard errors (in brackets) of the estimates are reported on the first line. RMSE values are given on the second line.

	Parameter	$\sigma_{11}$	$\sigma_{22}$	$\sigma_{12}$
	True value	200.00	300.00	150.00
$\delta = q_{0.25}$				
	RT	207.38 (7.24) <b>10.33</b>	299.57 (11.98) <b>11.98</b>	138.27 (7.98) <b>14.18</b>
	LT	199.91 (18.79) 18.77	300.53 (31.43) 31.41	150.21 (18.07) 18.07
	PRS	201.55 (19.56) 19.60	305.64 (32.64) 33.09	152.30 (20.97) 21.08
$\delta = q_{0.50}$				
	RT	225.64 (6.59) 26.47	309.70 (10.52) <b>14.30</b>	122.79 (7.06) 28.11
	LT	200.16 (19.56) <b>19.56</b>	300.98 (32.61) 32.61	150.39 (18.65) <b>18.65</b>
	PRS	202.43 (23.68) 23.78	307.70 (39.42) 40.12	153.23 (25.76) 25.93
$\delta = q_{1.00}$				
	RT	315.27 (7.45) 115.51	397.33 (10.25) 97.87	120.22 (7.19) 30.63
	LT	200.56 (22.75) <b>22.75</b>	302.21 (37.72) <b>37.75</b>	150.96 (21.70) <b>21.70</b>
	PRS	204.09 (31.00) 31.24	311.48 (52.11) 53.31	155.00 (33.83) 34.17

Table 3: Simulation results for a grid of 49 sites with spatial lag  $k = 10$ . For each estimation approach (RT, LT, PRS), the mean and the standard errors (in brackets) of the estimates are reported on the first line. RMSE values are given on the second line.

Table 4: Bias values for the RT and LT estimates using all pairs and three different threshold values, namely set to the 0.9, 0.95 and 0.98 empirical quantile. The spatial lags  $k$  are taken equal to  $k = 5$  and  $k = 10$ .

		$k = 5$			$k = 10$		
		0.9	0.95	0.98	0.9	0.95	0.98
$\sigma_{11}$	RT	50.48	30.65	16.32	226.47	167.81	115.27
	LT	0.80	1.55	2.20	0.06	-0.04	0.56
$\sigma_{22}$	RT	40.03	19.27	6.31	222.36	155.91	97.33
	LT	0.99	1.89	3.47	0.18	0.26	2.21
$\sigma_{12}$	RT	-16.48	-18.61	-16.9	-8.3	-20.17	-29.78
	LT	0.65	1.33	2.31	0.10	0.07	0.96

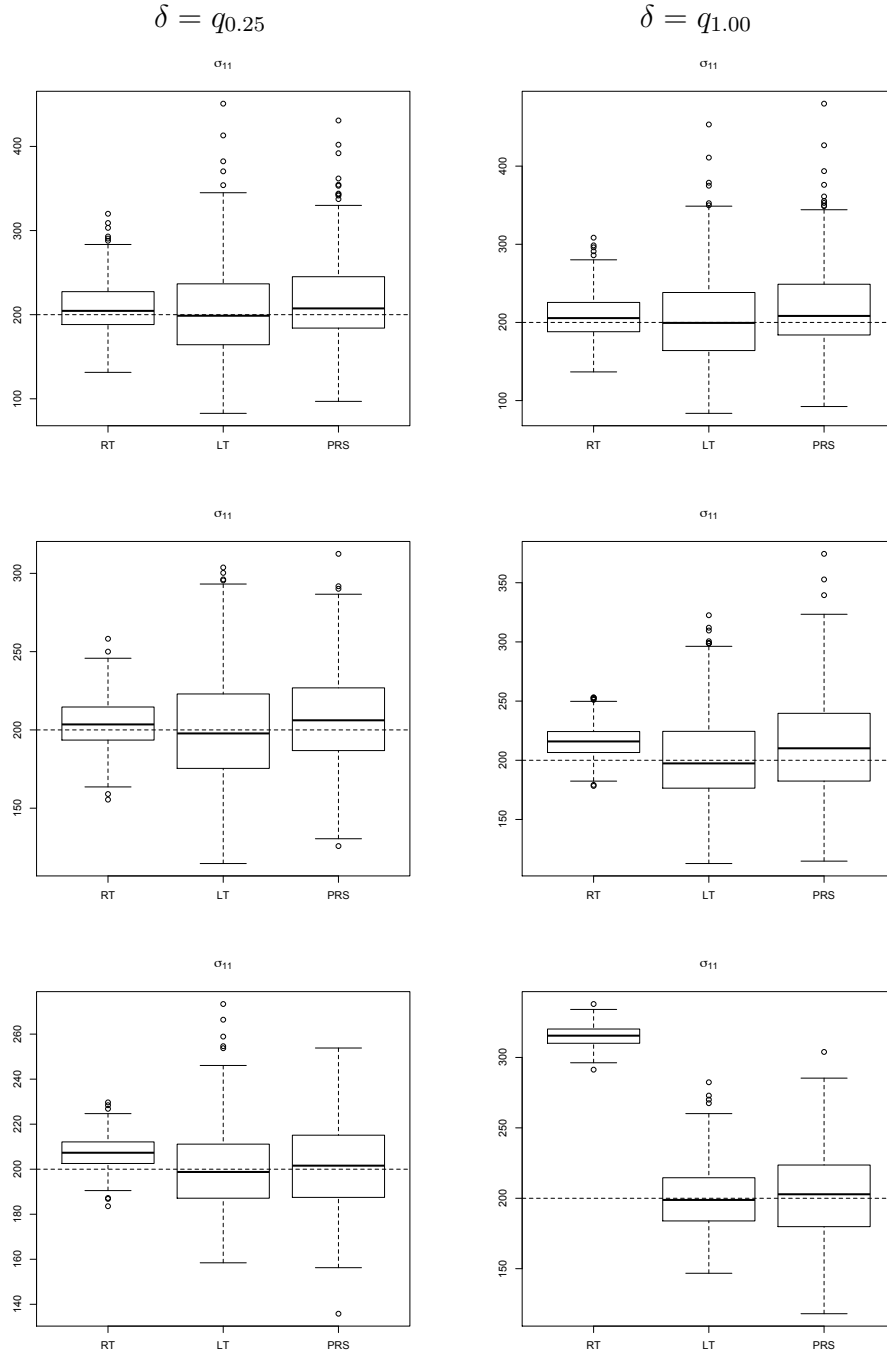


Figure 2: Box-plots of the estimates of  $\sigma_{11}$ . The true value for  $\sigma_{11}$  is indicated by a dotted line. Each row in the panel indicates the results over a grid with lag  $k = 1, 5, 10$ , respectively. In the first (resp. second) column of the panel the composite likelihood weights correspond to  $\delta = q_{0.25}$  (resp.  $\delta = q_{1.00}$ ).

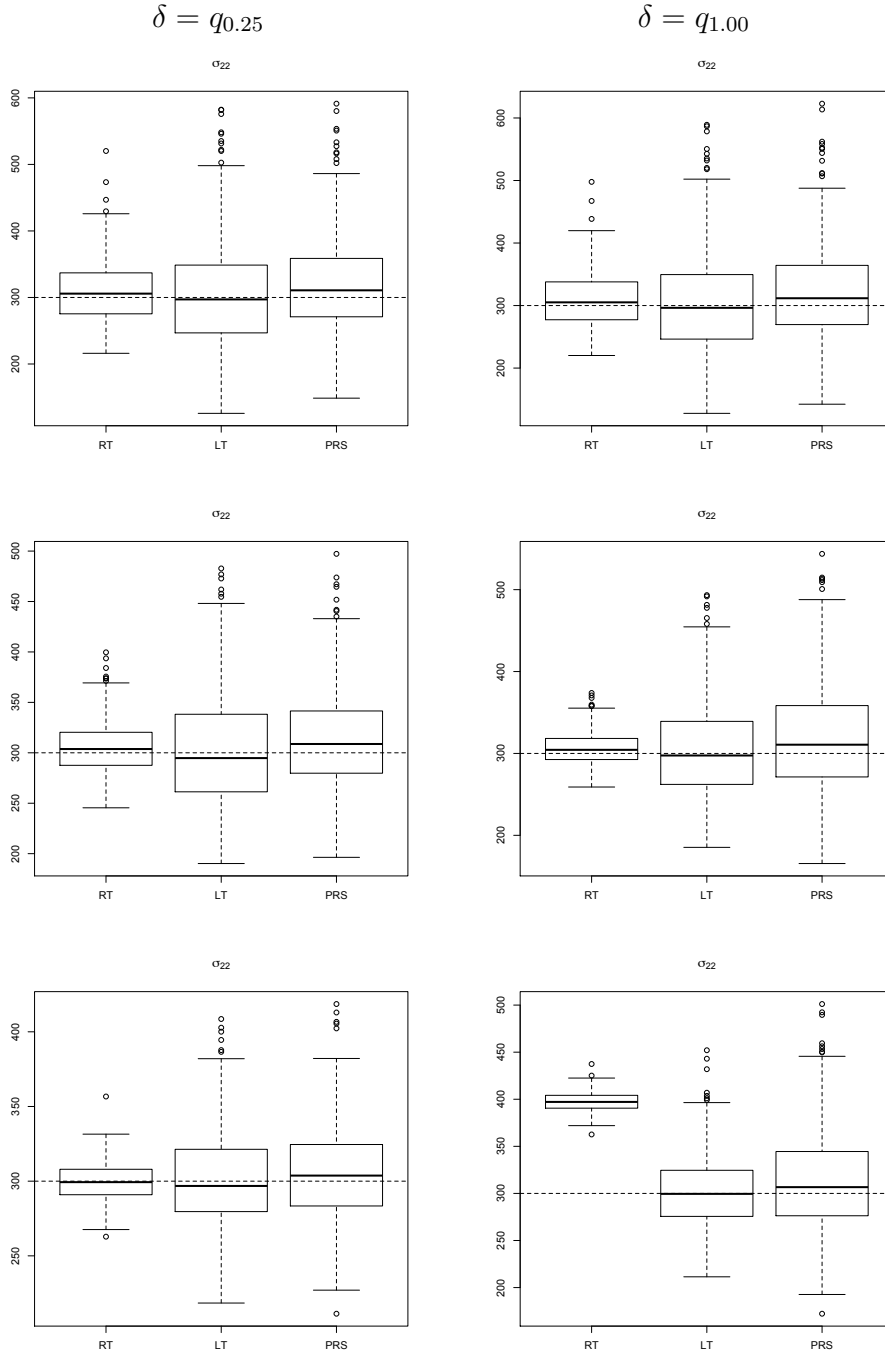


Figure 3: Box-plots of the estimates of  $\sigma_{22}$ . The true value for  $\sigma_{22}$  is indicated by a dotted line. Each row in the panel indicates the results over a grid with lag  $k = 1, 5, 10$ , respectively. In the first (resp. second) column of the panel the composite likelihood weights correspond to  $\delta = q_{0.25}$  (resp.  $\delta = q_{1.00}$ ).

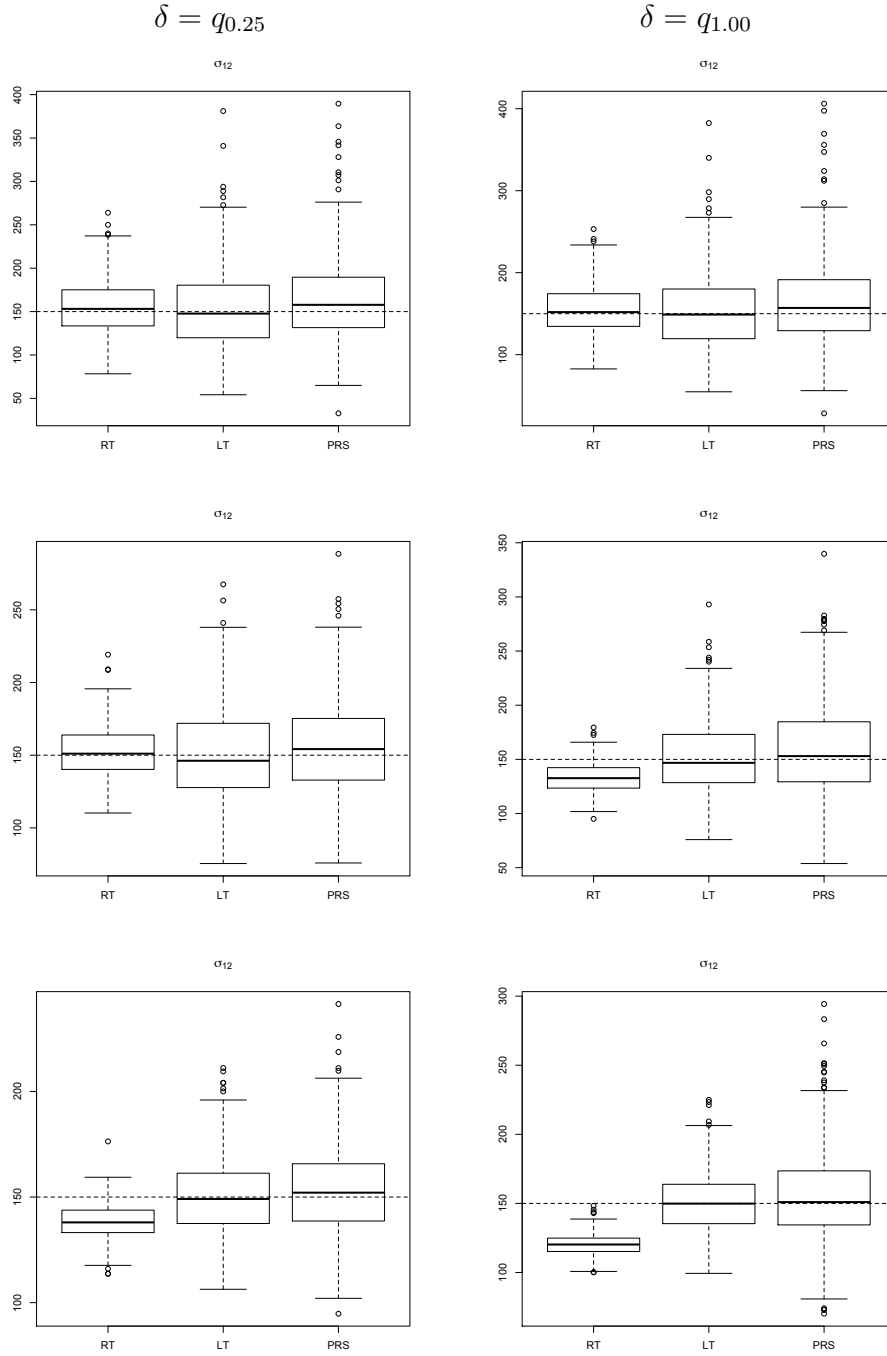


Figure 4: Box-plots of the estimates of  $\sigma_{12}$ . The true value for  $\sigma_{12}$  is indicated by a dotted line. Each row in the panel indicates the results over a grid with lag  $k = 1, 5, 10$ , respectively. In the first (resp. second) column of the panel the composite likelihood weights correspond to  $\delta = q_{0.25}$  (resp.  $\delta = q_{1.00}$ ).